Unsupervised Long-Term Routine Modelling using Dynamic Bayesian Networks

Yangdi Xu, David Bull and Dima Damen
Visual Information Laboratory, University of Bristol, Bristol, UK
Email: {Yangdi.Xu, Dave.Bull, Dima.Damen}@bristol.ac.uk

***Abstract*—Routine can be defined as the frequent and regular activity patterns over a specified timescale (e.g. daily/weekly routine). In this work, we capture routine patterns for a single person from long-term visual data using a Dynamic Bayesian Network (DBN). Assuming a person always performs purposeful activities at corresponding locations; spatial, pose and time-of-day information are used as sources of input for routine modelling. We assess variations of the independence assumptions within the DBN model among selected features. Unlike traditional models that are supervisedly trained, we automatically select the number of hidden states for fully unsupervised discovery of a single person's indoor routine. We emphasize unsupervised learning as it is practically unrealistic to obtain ground-truth labels for long term behaviours.**

The datasets used in this work are long term recordings of non-scripted activities in their native environments, each lasting for six days. The first captures the routine of three individuals in an office kitchen; the second is recorded in a residential kitchen. We evaluate the routine by comparing to ground-truth when present, using exhaustive search to relate discovered patterns to ground-truth ones. We also propose a graphical visualisation to represent and qualitatively evaluate the discovered routine.

I. Introduction

Research for elderly care assistance in smart home environments has become a popular research topic in recent years. For example, [1] gives an overall review of smart home technology; and [2] explains the effectiveness of smart home technology in social and health care support. However, smart home automation is often hindered by the challenges of analysing long term data and defining semantic labels for daily activities. In this paper, we focus on modelling a person's routine as a concise and commonly understood representation of the person's behaviour. *Routine* is accepted by the general public as the common, regular or standard course of activity patterns. Our work aims to model a person's *routine* by analysing relatively long term data using a single RGB-D camera in an indoor environment. Such modelling can facilitate automatic healthcare for occupants by detecting changes and out-of-routine activities.

Due to inter-personal differences and weak association to semantic labels, routine modelling is difficult to achieve in a supervised manner practically. While unsupervised deterministic models of routine such as [3] have been proposed, they use heuristic thresholds that make them difficult to generalise beyond the used datasets. In attempting to solve the problem, we were inspired by research on activity recognition using probabilistic graphical models. Although standard Hidden Markov Model (HMM) is commonly used for activity recognition, it is not sophisticated enough for activity pattern modelling. The Hierarchical Hidden Markov model (HHMM) is proven to be useful for supervised activity patterns modelling. Its graphical model consists of a *terminate state* which requires labeled information to determine when a certain activity stops. This indicates supervised learning for the model. Thus we propose to model the routine using a Dynamic Bayesian Network (DBN). Our DBN avoids using heuristic thresholds and incorporates routine-independent priors to improve performance.

Importantly, we emphasise the relationship between activity patterns and time - in routine modelling we are not only interested in what happens, but also in when does it happen. For example, 'morning coffee' and 'evening coffee' share the same activity pattern 'making coffee', but at different times of day. Therefore, we employ the concept of *time envelops* to encode durations relative to the start of the day where activities are common or frequent. These are discovered from time of day, and each frequent activity pattern is linked to at least one envelope. The effectiveness of integrating time envelopes in routine modelling is experimentally shown. Additionally, both location and pose are used as a combined feature in activity representation.

Related works are discussed in Sec. II. The method is detailed in Sec. III including the selection of model and parameters. The description of the dataset is in Sec. IV. Results are shown in Sec. IV-A including a visualisation approach to modelled routines in Sec. IV-B.

II. Related Work

Research for elderly care assistance in smart home environment has become a popular topic in recent years. Understanding the long term daily routines for people can better help to take care of their wellbeing. Researchers in the field of sensor networks frequently use smart meters and on-body sensors to model daily routines in indoor environments. As routine is closely related to different time periods of the day, [4] proposed using a combination of motion (PIR) and door sensors to discover the time intervals of certain events. [5] proposed a method to extract spatio-temporal activity models using a wireless sensor network inside a home. A device triggered by the moving person is used to generate labelled activities. Eventually, the daily routine model of a person is built by finding the most probable, network-level, sequences of nodes, sensing features, namely location, time and duration. [6] also modelled activities in a smart home environment. They introduce an unsupervised method to firstly discover the frequent and repeatable sensor events, then cluster them into groups of activities to recognise each activity class.

Although using visual data has the advantage of being remote, able to observe how the action was performed and less hardware driven, most research work in Computer Vision is related to activity recognition and few have attempted routine modelling. However, since modelling activities is part of the routine modelling process, it is useful to discuss relevant approaches to activity recognition. [7] proposed deterministic models to discover activities using visual data with unstructured scenes. They used an unsupervised hierarchical activity model which combines global and local motion features to build activity events at different resolutions. [8] used Markov Clustering Topic Model (MCTM) to address the problem of unsupervised mining of multi-object spatio-temporal behaviours in crowded and complex public scenes. The model has a two layer structure and is learned offline from unlabelled training data with Gibbs sampling. But in our case, the time envelopes can hardly been incorporated into the topic model.

A statistical model does not require any heuristic thresholds and is able to explain results as well as calculate confidence levels. [9] introduced multi-scale statistical model for behaviour representation called an Abstract Hidden Markov Model (AHMM). It is similar to a Hierarchical Hidden Markov Model (HHMM) but the termination of high level behaviour has a direct connection with the lower level behaviour in a hierarchical model. The spatial and trajectory information are used as their primary features. Brand *et al* [10] use coupled HMMs for supervised complex activity recognition. Two HMMs are coupled by introducing conditional dependencies between their state variables. The result shows that coupled HMM are typically good at modelling activities that do not strictly obey the Markovian assumption, such as interacting processes. It is far less sensitive to initial conditions than conventional HMMs. HHMM was used in [11] for monitoring behaviours. Rather than learning all parameters of HHMM at once, they proposed a two-phase learning algorithm to learn the parameters of primitive behaviour first. [10] coupled two HMMs by introducing conditional dependencies between their state variables for supervised complex activity recognition. The results show that coupled HMMs are capable of modelling activities that do not strictly obey the Markovian assumption, such as interacting processes. [12] developed a Dynamic Multi-Linked Hidden Markov Model (DML-HMM) to discover temporal and causal correlations among discrete events for behaviour interpretation. Their results show that DML-HMM has better performance in modelling activities in a noisy and cluttered scene compared to other DBN models such as Multi-Observation HMM and Coupled HMM. [13] proposed a novel on-line forward-backward relevance algorithm on top of an off-line Multi-Observation HMM to model the outdoor activities on surveillance footage. All the works above are achieved in a supervised manner. The number of states for each hidden node is determined manually by using the knowledge of the dataset.

The terminology *routine* has a finer granularity than that of activities, as each *routine* should contain one or more frequent reoccurring activities in different time periods. [3] provided a clear definition of routine. In their proposed method, the frequent transitions in spatial and pose features are used to represent activity patterns. They are discovered using a combination of top-down and bottom-up hierarchical method to find frequent activity patterns. However, [3] uses a number of heuristic thresholds which makes it difficult to generalise. Our work focuses on modelling *routine* patterns using an unsupervised probabilistic approach. To tackle the problem, we use a general Dynamic Bayesian Network (DBN). DBNs capture temporal causality and allow modelling dependencies within each time step. To generalise to any routine, we do not make assumptions about interacting processes.

Our work focuses on modelling *routine* patterns in an indoor environment using an unsupervised general Dynamic Bayesian Network (DBN) model. The contributions are presented as: (i) Our DBN is trained using unlabelled data and the number of hidden states for each node within the DBN is determined automatically. (iii) Time envelopes are included based on the inseparable connection between time and activities in routines. (iii) We systematically assess the independence assumptions in the DBN model, (iv) We test on datasets with non-scripted behaviours in their natural environments to model routine for a single person, and (V) A graphical visualization is proposed to qualitatively analyse and communicate routines without ground-truth.

III. METHOD

A Dynamic Bayesian Network (DBN) is a probabilistic model that represents a set of random variables and their dependencies over adjacent time steps, with two types of nodes: hidden and observed. Each node is associated with a probabilistic function that takes the variables of parent nodes as input, and models the probability distribution function for the child node. In addition, it is natural that the person will interact with certain objects to perform activity patterns in an indoor environment. Thus, the frequent visited locations are considered as 'hot spots'. Different human poses at different 'hot spot' and time are used to represent indoor routines. Therefore, we consider three observed nodes at each time step: 3D location of the person in the environment, O_t^L, the body pose O_t^H and the corresponding time of day O_t^E. Each observed node is associated with a hidden node that attempts to model discrete frequent locations L_t, poses H_t and time envelopes E_t respectively. The notion of *time envelopes* is inspired by routine modelling in sensor networks [4], where it refers to the loose time period that always includes the routine event. Here we model these time envelopes as latent variables (i.e hidden nodes) that are discovered from time-of-day information using clustering.

Figure 1 shows three DBNs with different independent assumptions, with Figure 1(A) shows our proposed model. This is based on the assumption that the spatial, pose and time features are only dependent given an activity. Thus, there are no dependencies among the first layer hidden nodes. In (B), we encode dependencies in the first layer, where we assume that given the time of day, the person is more likely to visit certain locations, which may lead to performing certain poses. In (C), we investigate the need for incorporating time-of-day in the model, by excluding its observed and hidden nodes. The proposed DBN (A) and its variations (B), (C) will be evaluated in Sec IV-A.

To model the routine using the proposed DBN in Figure 1(A), for example, we need to sequentially learn four aspects of the model:

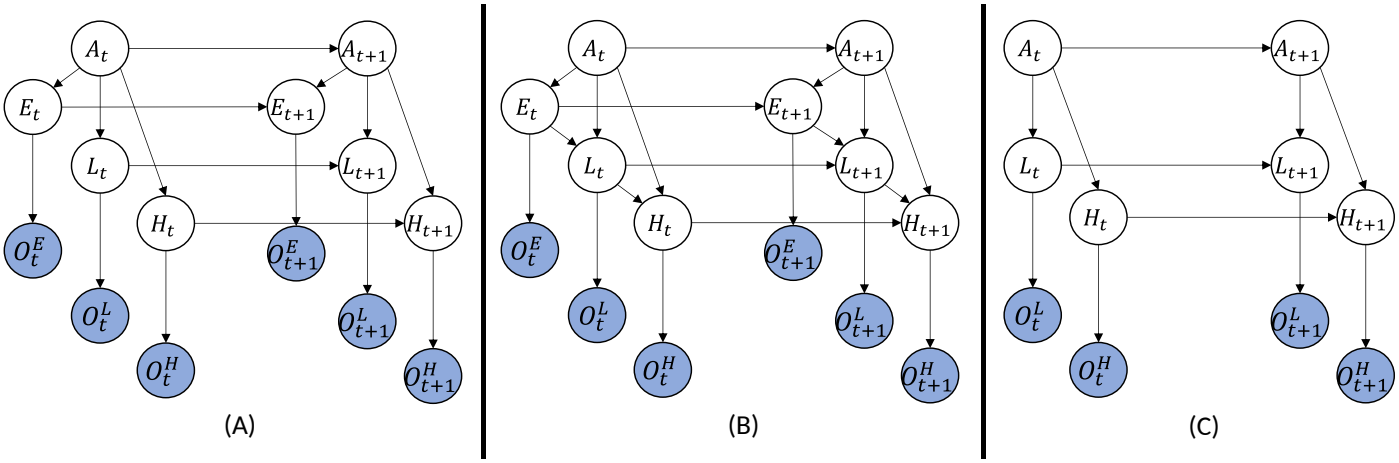

Fig. 1. Three DBN models: (A) Proposed Model; (B) Variation 1: Dependencies in first layer; (C) Variation 2: Without Time Envelope.

1) The conditional probabilities between the observed and hidden nodes is estimated using a Gaussian Mixture Model, with the number of Gaussians assumed to be equal to the number of states in the hidden node; namely $P(O_t^H|H_t), P(O_t^L|L_t)$, and $P(O_t^E|E_t)$.

2) The conditional probabilities between hidden nodes within the same time step; namely $P(L_t|A_t)$, $P(H_t|A_t)$, and $P(E_t|A_t)$.

3) The conditional probabilities between consecutive time steps; namely $P(A_{t+1}|A_t)$, $P(L_{t+1}|L_t, A_{t+1})$, $P(H_{t+1}|H_t, A_{t+1})$, $P(E_{t+1}|E_t, A_{t+1})$.

4) Prior probability of activities $I(A_t)$ which we assume to be uniform.

Before the graphical model can be trained, we must decide on the number of states for each hidden node. For the first layer of hidden nodes [E_t,L_t,H_t] we use K-mean clustering to group data points as locations of interest, pose clusters and time envelopes. We use the Elbow Test [14] to estimate the optimal number of clusters, and thus the number of states, avoiding heuristic thresholds.

$$R(k) = \Big(\sum_{i=1}^{k}\sum_{j=1}^{l_i}|(Y_{ij} - \overline{Y}_i)|\Big) / \Big(\sum_{i=1}^{k}\sum_{j,i\neq j}^{k}|(\overline{Y}_i - \overline{Y}_j)|\Big) \quad (1)$$

In equation 1, k is the number of clusters, l_i is the total number of points in cluster $1 \leq i \leq k$, and Y represents the data points. For each k, $R(k)$ calculates the sum of intra cluster distances over the sum of inter cluster distances. The optimal number of clusters is then determined where R converges. The condition of the optimum value of k is shown equation 3,

$$g(k) = R(k+1) - R(k) \quad (2)$$

$$\hat{k} = \min k \quad where \quad |g(k)| - |g(k+1)| \geq 0 \quad and \quad g(k+1) \geq 0 \quad (3)$$

The number of states for the second layer of hidden nodes, that is of the Activity node A_t, which we refer to as N, is determined by the likelihood value of the model. Once the optimum numbers of states k_L, k_H and k_E are determined by the 'Elbow Test', we train the DBN across a certain range of N values. During the EM optimisation, the log-likelihood value is calculated for each model. As the value indicates how well the system has fitted the data, it can be used to choose the optimum value of N for each dataset.

We use Murphy's Bayesian Network toolbox [15] for model implementation. Since our proposed model is a dynamic model with full observations, we use the *junction tree inference engine* together with *Expectation Maximisation (EM)* for inferencing and parameter tuning. During training, the transition matrices are estimated. We randomly initialised these parameters and apply EM on top of the forward-backward inference. Theoretically, the DBN should converge to an optimal result. However, it often falls into a local optimum due to the large number of parameters to tune. One solution is to use simulated annealing for approximating the global optimum. The drawback is the lengthened computation time. Instead, we introduce *dataset-independent* priors to the transition matrices based on the assumption that the state is more likely to remain the same during transitions between time t and $t+1$. For example, for transition matrix $P(A_{t+1}|A_t)$, we give a high probability along the diagonal and significantly smaller probability ϵ elsewhere. This prior is understandably closer to the optimal solution than a random initialisation. The assumption only works on the transition matrices connecting hidden nodes between t and $t+1$. The remaining conditional probabilities are randomly initialised.

For all datasets (Sec. IV), the OpenNI library is used to extract the 3D location, as well as silhouettes using background subtraction [16]. The spatial observation node O_t^L represents the 3D coordinate of the center of body. The pose observation node O_t^H represents silhouettes scaled to 100×60 then reduced in dimension using Principal Component Analysis (PCA) to the first 80 components. Despite recording using an RGB-D sensor, although the quality of some of the estimated poses from non-frontal viewpoints is poor and cannot be reliably used, the example is shown in Fig 2; we managed to obtain

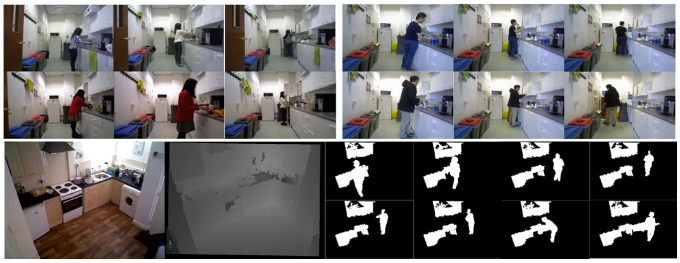

Fig. 2. Top: office kitchen recordings. Bottom Left: The RGB and depth image of the residential kitchen. Bottom: Right: Example recorded silhouette images of residential kitchen.

clean silhouettes using our own customised background subtraction technique(result shown in sec IV-B). Finally, the time-of-day observation node O_t^E is the time stamp in milliseconds relative to the start of the day. This could be changed should a different scale of routine be considered (e.g. weekly routine, monthly routine).

IV. Datasets and Results

Collecting visual data for routine modelling is a challenging task. Unlike action and activity recognition, datasets that are viable for routine modelling need to exhibit three characteristics. First, the recorded activities need to appear frequently to be considered as routines. Second, the dataset should be a long-term recording. By long-term we do not refer to the length of each individual recorded sequence, but to the range of time that the dataset covers. Activities can only be considered as routines if they occur frequently in period of days, weeks or even months. Therefore, those datasets that only focus on action or activity instances cannot be used for routine modelling. Finally, the recording should be non-scripted - the participant is free to act according to his/her own behavioural habits. Up to our knowledge, no publicly available long-term dataset with visual or depth sensors are available for routine analysis. For example, the TUM Kitchen Dataset [17] is used for kitchen activity analysis, but is actually a simulated lab environment. The recording is not long-term as their research focus is on analysing multiple individuals performing the same task. The morning dataset from [18] and Watch-n-Patch dataset [19] both contain no repetitive activities per person, and it is mostly scripted, which do not meet the requirement for routine discovery.

We thus present two datasets for routine modelling, captured in kitchens as activities in the kitchen tend to be goal oriented. The two datasets are from office and residential kitchens respectively. The datasets are summarised in table I, and the details are described next. Both dataset will be public available in the future.

Office Kitchen Dataset: The office kitchen dataset contains four sets of recordings, from one viewpoint, shown in figure 2. Each set captures all kitchen activities of one individual over a period of 6 days using a single RGB-D sensor. We refer to each visit to the kitchen as a 'sequence'. The dataset captures three individuals, one recorded twice with the recordings being fifteen months apart. All activities performed are non-scripted in order to obtain a natural behavior pattern. In this dataset, the skeleton data, RGB and depth images are available to use. Figure 3 shows an example of the ground-truth patterns from three days captured in office kitchen 1. Each timeline represents a single day. As seen from the figure, kitchen activities take place during the morning and afternoon frequently, and less frequently in the early evening. Even within the morning routine, the starting time varied between 9:49A.M. and 10:23A.M, which highlights the need for encoding time envelopes rather than exact times. The sub-activities within the same time envelope are similar, but in no way identical.

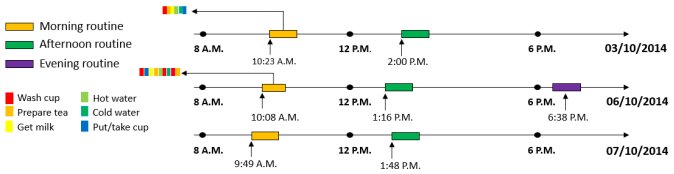

Fig. 3. Office Kitchen 1 routine patterns on time line of three working days.

Residential Kitchen Dataset: The participant in this dataset lived in a sensor-equipped house for one week. The dataset only captures the participant's morning activities (6am - 12pm). This represents the most realistic dataset on routine in the vision community to date. Data is recorded using an RGB-D sensor in the kitchen. However, while this dataset presents huge opportunities, a practical obstacle is that the raw data captured only silhouettes and 3D bounding box information due to privacy concerns. Unlike office kitchen recordings, the average length of the sequence is significantly longer and the variance of recording length is also higher. For example, cooking a proper meal at home may take more than half an hour whereas making a cup of tea only takes several minutes. Figure 2 shows the layout of the residential kitchen in both RGB and depth, as well as sample recordings.

Note that while ground-truth labels could be provided for the Office Kitchen dataset by watching the videos, this is not possible for the Residential Kitchen recording. We address this in Sec. IV-B.

A. Experiments and Results

We first report results on the Office Kitchen dataset for which we have ground-truth labels. However, in unsupervised modelling, quantitative evaluation is challenging, even when ground-truth is available. Figure 4 shows a common problem, where the ground-truth consists of four semantic labels for a recorded sequence. The discovered activity patterns from the DBN do not have semantic labels associated with them, as they are automatically discovered. They thus are not expected to match the ground-truth labels perfectly. We use an exhaustive search method to map the discovered patterns to the corresponding ground-truth labels to calculate accuracy for evaluation purposes. We match each ground truth label to the modelled routine pattern that has the highest overlapping frames. This one-to-one match could result in discarding some frames as false positive. After this assignment is achieved, discovered patterns are labeled from which the *accuracy* can be calculated as the percentage of frames with a correct match to ground-truth.

Due to random initialisation, the accuracy might not be exactly the same for every optimisation iteration. We thus report

TABLE I. INFORMATION SUMMARY OF DATASETS AND RECORDED SEQUENCES; PID: PARTICIPANT ID, #SEQ.: NUMBER OF SEQUENCES RECORDED, L_{max}: MAX LENGTH, (S): IN SECONDS

Dataset & Recordings		MM/YYYY	PID	Gender	#Seq.	L_{max}(s)	L_{min}(s)	L_{mean}(s)	L_{sd}(s)
D1	Office Kitchen 1	10/2014	P1	F	16	176	39	98	42
	Office Kitchen 2	01/2016			17	152	37	90	33
	Office Kitchen 3	02/2016	P2	M	15	100	28	69	23
	Office Kitchen 4	02/2016	P3	M	15	122	23	56	26
D2	Residential Kitchen	04/2016	P4	M	19	2520	65	502	596

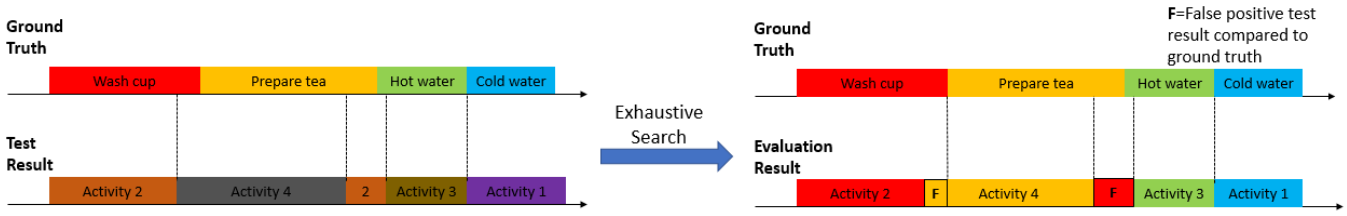

Fig. 4. An example of unsupervised results interpretation. Using exhaustive search for evaluation.

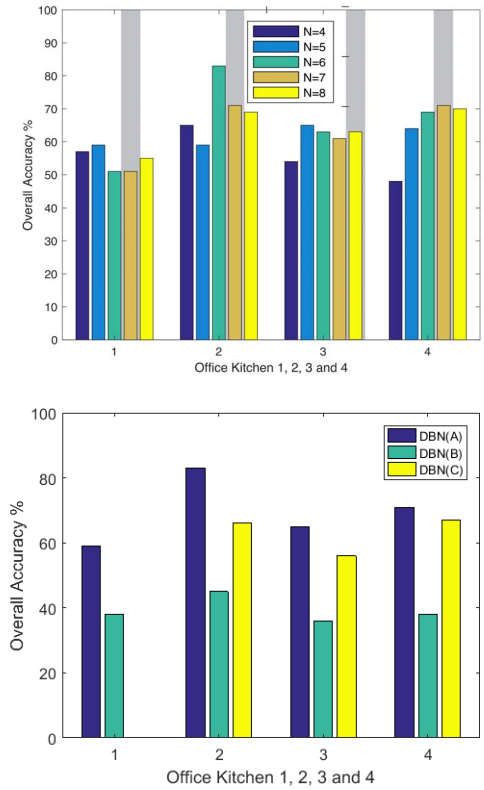

Fig. 5. Top: Accuracy for number of activities N; Gray shading indicates chosen values by the proposed unsupervised model. Bottom: Accuracy for DBN models.

the accuracy from the average of 15 iterations. In figure 5, we show the accuracy as the number of discovered activities N changes. The optimal N resulting from automatically selecting the number of states for all nodes in the DBN is shaded in gray for each case. For recording 4, the optimal N matches the highest accuracy. The biggest drop in accuracy is in Office Kitchen 2, for which we show an analysis of the optimal number of state combination of K_L and K_H in table II. In this case, the chosen number of states by the Elbow Test

TABLE II. ACCURACY FOR NUMBER OF STATES FOR OFFICE KITCHEN 2

k_L/k_H	4	5	6	7	8
3	0.64	0.62	0.66	0.59	0.62
4	0.64	0.62	0.66	0.56	0.57
5	0.62	**0.80**	0.67	0.65	0.57
6	0.56	0.65	0.61	0.63	0.64
7	0.69	**0.80**	0.60	0.51	0.60
8	0.55	0.63	0.71	0.74	0.58

TABLE III. AUTOMATIC SELECTION OF OPTIMAL NUMBER OF STATE USING UNSUPERVISED APPROACH. N IS THE NUMBER OF ACTIVITY PATTERN, k_L IS THE NUMBER OF LOCATION 'HOT SPOT', k_H IS THE NUMBER OF POSES, k_E IS THE NUMBER OF TIME ENVELOPE.

	N	k_L	k_H	k_E
Office Kitchen 1	7	6	5	3
Office Kitchen 2	7	5	6	3
Office Kitchen 3	8	4	7	3
Office Kitchen 4	7	5	5	3
Residential Kitchen	4	4	5	2

clearly affects the accuracy. The complete results on optimal number of states for each hidden node is shown in table III. Although the model performance of auto parameter selection is not the highest when compared with ground-truth, they fits DBN model the best using the current data. Since the ground-truth is manually labelled and is not been used in learning the DBN model, the labels may not be the best representation of the data. Thus, there is a slight decrease in performance when we compare our result with the manual ground-truth.

Figure 5 (bottom) also displays the result comparison between the three proposed DBNs (Sec III). It shows that by using the independent proposed model with time information - DBN(A), our routine modelling achieves higher accuracy than the two variations on all recordings. This is because we believe that pose/location/time will remain independent unless they are related by a regular activity pattern. When pairwise dependencies are incorporated in DBN(B), confusion may be added due to the a force connection between similar poses and location hot spots. In fact (B) may perform better if a person

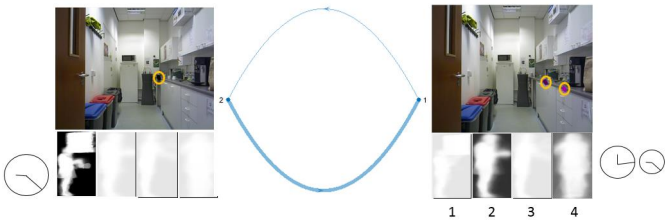

Fig. 6. An example graphical visualization with associated correlations. Descriptively, this sample routine shows the person around 9:20 going first to the cupboard before approaching the coffee machine

performs distinct poses at distinct locations and times - which in practice does not apply as it requires larger and more spread out indoor space. The result also highlights the importance of modelling time envelopes for routine discovery, as DBN(A) outperforms DBN(C) in every case. The biggest difference is shown for office kitchen 2, where the accuracy drops from 83% to 66% when time envelopes are excluded, which is a 20.1% drop in performance. Note the missing value for Office Kitchen 1 on DBN(C) as more than one transition matrix become singular during optimisation and stops the model from converging.

B. Graphical Visualisation

As no ground truth is available for the Residential Kitchen dataset, and similarly for any long-term routine data, we introduce a graphical visualisation that demonstrates all the elements of the discovered routine. The visualisation captures temporal associations between frequent activity patterns, frequent locations, frequent poses and their corresponding temporal envelopes. We believe that such visualisation could be appropriate for qualitatively evaluating our unsupervised routine modelling.

Figure 6 shows an example of the proposed graphical visualisation. It consists of four main parts: the activity pattern graph (middle), associated locations, poses and a 'clock' to display the time envelope. In the activity pattern graph, the width of the arrow indicates the degree of temporal correlation between different patterns, while the arrow indicates which pattern precedes the other temporally. The location of the person is plotted as a Gaussian distribution. The darker colour, highlighted by orange circles, indicates the person is more likely to appear at certain locations with the given activity pattern. In the sample visualization in Figure 6, the location graph shows a very strong correlation of activity pattern (left) with the worktop in the kitchen, as well as high correlation between pattern (right) and both the sink and the coffee machine. Each activity pattern is also associated with certain pose clusters - visualised using the cluster mean of the silhouettes. The transparency is related to the probability of the pose occurring. The 'clock' visualization shows the mean time of the associated time envelopes, with size indicating probability.

Figure 7 is the graphical visualisation for the residential kitchen. It illustrates time envelopes of *8:20* and *10:20*. It also shows a strong correlation (shaded in red) among three different patterns at $N = 4$. A similar correlation is found at $N = 6$ with some remodelling in spatial and pose patterns. The area shaded in purple shows a clear split in spatial patterns between $N = 6$ and $N = 7$ with few changes in pose. The blue area shows that two activity patterns share the same spatial and pose combination, with the only difference being the probability associated with each location. Visually, one could conclude that the routine is sufficiently presented at $N = 4$, which is in fact the selected optimal N. At $N = 4$, the discovered routine for this dataset are: *use microwave* to prepare meal(top), *wash dishes*(right), *boil water*(left) and *make tea*(bottom). Be aware that since we don't have ground-truth, the above semantic routine for this dataset is obtained by intuitively understanding of the silhouette video, which is less accurate than the conclusions we can make about the Office Kitchen dataset.

Figure 8 shows routine patterns for Office Kitchen 2. Although manual ground-truth is available, we present it in a graphical form for consistency and comparison purposes. The figure shows three discovered time envelopes: *10:20*, *14:10* and *16:55*. When $N = 6$, we see a strong correlation in the red shaded area which transit from bottom left to top right. It shows the temporal links between the worktop, fridge, boiler and sink with the corresponding pose transitions. Compared with ground-truth, the discovered routine is: *get milk from fridge*,*wash cup* and *make beverages* in the time envelope around *10:20* and *14:10*. While we observe similar patterns and correlations at $N = 7$, both the spatial and pose patterns in the red shaded areas have been remodelled into two different patterns. Also comparing $N = 4$ to $N = 5$, the green shaded area shows a clear split in spatial, pose and time. It leads to a separation of different poses resulting in a better modelling of activity patterns. This hopefully demonstrates that by using different number of N, the activity patterns are explained at different levels of granularity.

V. Conclusion and Future Work

In this work, we proposed a Dynamic Bayesian Network (DBN) for modelling daily routines of a single person in indoor environments. It uses spatial and pose features and incorporates the concept of time envelopes. Two additional variations of the model are evaluated based on the different dependency assumptions among the features. All models are learned in an unsupervised manner and dataset-independent initialisations. Results show that the proposed DBN which incorporates the knowledge of the time envelope achieves the highest accuracy. Finally, a graphical visualisation presents the discovered routine patterns and their corresponding correlations when ground-truth is not available.

Our future work will focus on discovering routine changes. The Office Kitchen dataset will be used as recording 1 and recording 2 are of the same person albeit 15 months apart. Visible routine changes can be observed in this dataset and could be automatically discovered. Further recordings could also be obtained.

Data Statement & Ack: Office Dataset and annotations are available on the project's webpage: https://www.cs.bris.ac.uk/~damen/Routine/. Residential dataset cannot be released for privacy reasons and is part of the SPHERE project. Y Xu has been partially supported by EPSRC-IRC project SPHERE (EP/KO31910/1). Work also supported by EPSRC LOCATE (EP/N033779/1).

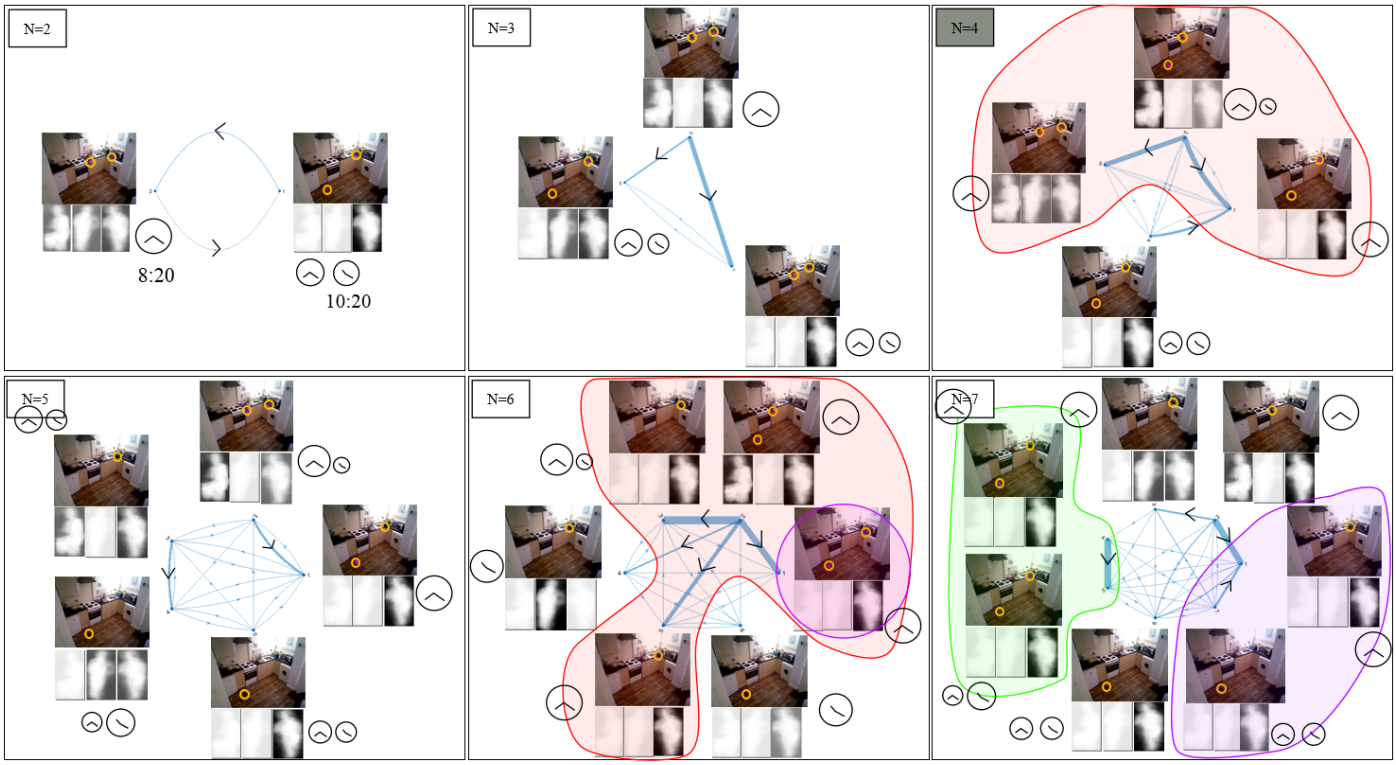

Fig. 7. Graphical visualization for residential dataset. It presents the correlations among location, pose and time for different number of activity pattern N. Shaded areas highlight related, typically split, patterns as N changes. The shaded optimal N in this case is $N = 4$.

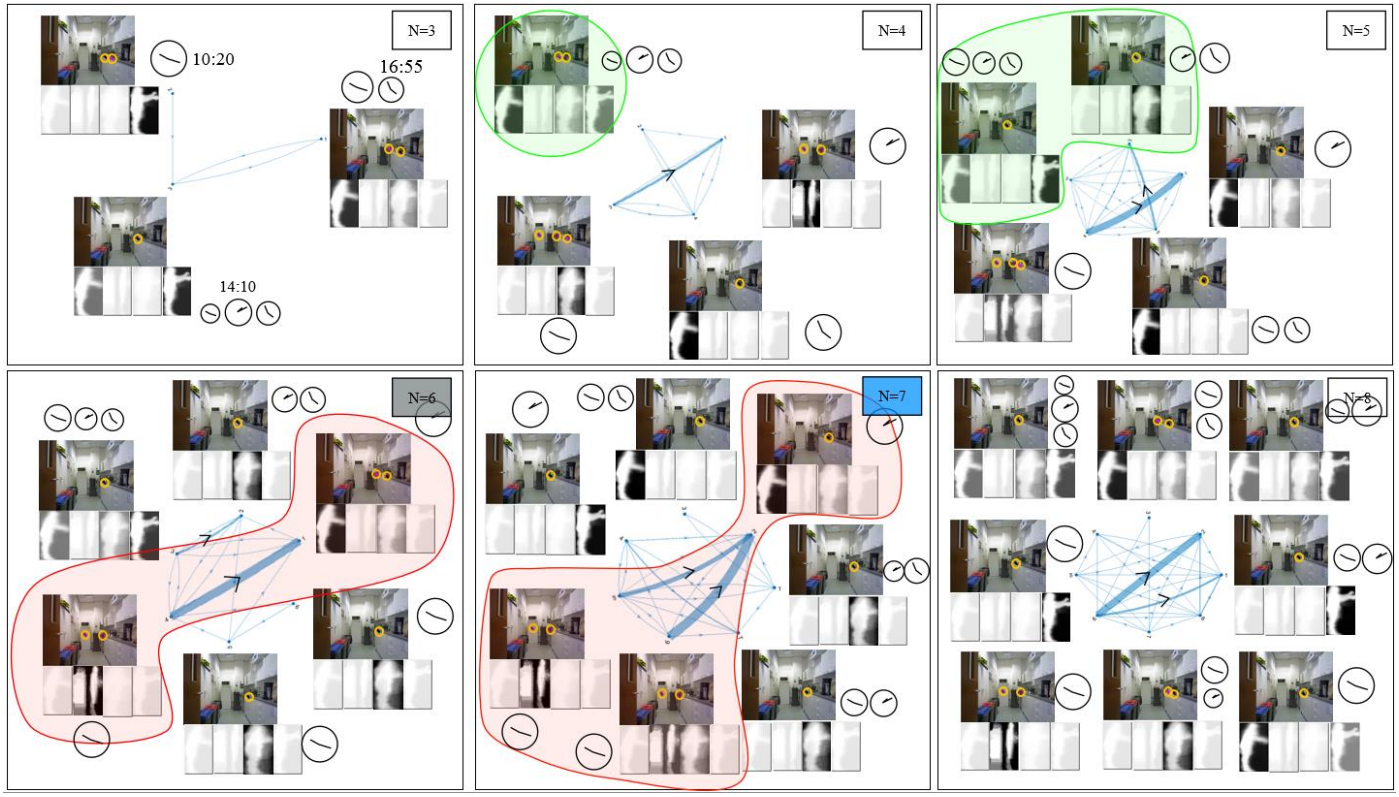

Fig. 8. Graphical visualization for Office Kitchen 2. $N = 6$ gives the highest accuracy when compared to ground-truth data. $N = 7$ is the optimal pattern number selected automatically by our method.

REFERENCES

[1] R. J. Robles and T.-h. Kim, "Applications, systems and methods in smart home technology: a review," *International Journal of Advance Science and Technology*, vol. 15, 2010.

[2] S. Martin, G. Kelly, W. G. Kernohan, B. McCreight, and C. Nugent, "Smart home technologies for health and social care support," *Cochrane Database Systematic Review*, vol. 4, 2008.

[3] X. Yangdi, B. Dave, and D. Dima, "Unsupervised daily routine modelling from a depth sensor using top-down and bottom-up hierarchies," in *Asian Conference on Pattern Recognition*. IEEE, 2015.

[4] J. Fang, A. Bamis, and A. Savvides, "Discovering routine events and their periods in sensor time series data," in *9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2010.

[5] D. Lymberopoulos, A. Bamis, and A. Savvides, "Extracting spatiotemporal human activity patterns in assisted living using a home sensor network," *Universal Access in the Information Society*, vol. 10, no. 2, pp. 125–138, 2011.

[6] P. Rashidi, D. Cook, L. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," *Knowledge and Data Engineering*, vol. 23, no. 4, pp. 527–539, 2011.

[7] S. Elloumi, S. Cosar, G. Pusiol, F. Bremond, and M. Thonnat, "Unsupervised discovery of human activities from long-time videos," *IET Computer Vision*, vol. 9, no. 4, pp. 522–530, 2015.

[8] T. Hospedales, S. Gong, and T. Xiang, "Video behaviour mining using a dynamic topic model," *International Journal of Computer Vision*, vol. 98, no. 3, pp. 303–323, 2012.

[9] S. Osentoski, V. Manfred, and S. Mahadevan, "Learning hierarchical models of activity," DTIC Document, Tech. Rep., 2005.

[10] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition," in *Computer Vision and Pattern Recognition*. IEEE, 1997, pp. 994–999.

[11] N. Nguyen and S. Venkatesh, "Discovery of activity structures using the hierarchical hidden markov model," in *British Machine Vision Conference*, 2005.

[12] T. Xiang and S. Gong, "Beyond tracking: Modelling activity and understanding behaviour," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 21–51, 2006.

[13] ——, "Activity based surveillance video content modelling," *Pattern Recognition*, vol. 41, no. 7, pp. 2309–2326, 2008.

[14] F. Can and E. A. Ozkarahan, "Concepts and effectiveness of the cover-coefficient-based clustering methodology for text databases," *ACM Transactions on Database Systems*, vol. 15, no. 4, pp. 483–517, 1990.

[15] K. Murphy, "Dynamic bayesian networks: representation, inference and learning," Ph.D. dissertation, University of California, Berkeley, 2002.

[16] open source, "OpenNI SDK," in *http://openni.ru/openni-sdk/*.

[17] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition," *International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS), in conjunction with International Conference on Computer Vision*, 2009.

[18] M. Karg and A. Kirsch, "Low cost activity recognition using depth cameras and context dependent spatial regions," in *International Conference on Autonomous agents and multi-agent systems*, 2014, pp. 1359–1360.

[19] C. Wu, J. Zhang, O. Sener, B. Selman, S. Savarese, and A. Saxena, "Watch-n-patch: Unsupervised learning of actions and relations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.